

## 特约评述

DOI: 10.12211/2096-8280.2023-083

## 文库构建与基因簇靶向筛选驱动的微生物天然产物高效发现

虞旭昶<sup>1,2</sup>, 吴辉<sup>2</sup>, 李雷<sup>1</sup>

(<sup>1</sup> 上海交通大学生命科学技术学院, 微生物代谢国家重点实验室, 上海 200240; <sup>2</sup> 华东理工大学生物工程学院, 生物反应器工程国家重点实验室, 上海 200237)

**摘要:** 微生物天然产物作为小分子药物的主要来源, 已被广泛应用于医药与农业等领域。随着全球抗生素耐药性与其他公共健康问题的加剧, 新结构、新靶点微生物天然产物发现迫在眉睫。大规模(宏)基因组测序揭示微生物蕴含了巨大的生物合成潜力, 相继催生了多种不同类型的天然产物挖掘策略。然而, 目前仍然缺乏将天然产物合成基因簇与编码产物快速关联的高效方案。近年来, (宏)基因组文库构建在获取批量天然产物合成基因簇方面展现出明显优势, 结合高效的基因簇靶向筛选方法, 显著加速了新结构天然产物系统发现。本文综述了三类基于(宏)基因组文库构建与靶向筛选驱动天然产物创新发现的策略, 主要从克隆载体类型、文库构建方式、基因簇靶向筛选方法等角度进行了阐述, 并对Cosmid/Fosmid文库、BAC/PAC文库、FAC/YAC文库等不同文库类型的优缺点及应用范围进行了对比, 最后对这些策略的发展前景进行了展望。未来, 基于文库构建与基因簇靶向筛选策略将极大驱动不同生境微生物来源的活性天然产物挖掘, 预期大量新靶点、新结构天然产物将不断涌现。

**关键词:** 微生物天然产物; (宏)基因组挖掘; 基因簇; 文库构建; 基因簇靶向筛选

中图分类号: Q819 文献标志码: A

## Library construction and targeted BGC screening for more efficient discovery of microbial natural products

YU Xuchang<sup>1,2</sup>, WU Hui<sup>2</sup>, LI Lei<sup>1</sup>

(<sup>1</sup> State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; <sup>2</sup> State key Laboratory of Bioreactor Engineering, School of Biological Engineering, East China University of Science and Technology, Shanghai 200237, China)

**Abstract:** Microbial natural products (NPs) are a major source for mining small molecule drugs, which have been widely used in medicine, agriculture, and other fields. Growing antimicrobial resistance and other public health problems necessitate the rapid discovery of microbial NPs with novel structures and bioactivities. With rapid advances

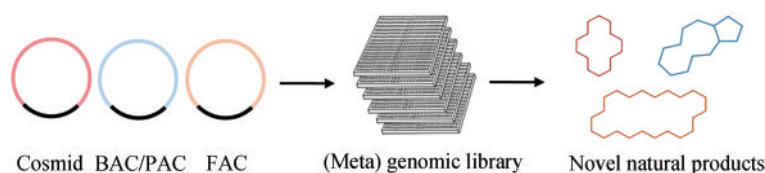
收稿日期: 2023-11-27 修回日期: 2024-01-26

基金项目: 国家重点研发计划(2023YFA0914200, 2023YFA0916200); 国家自然科学基金面上项目(32370070)

引用本文: 虞旭昶, 吴辉, 李雷. 文库构建与基因簇靶向筛选驱动的微生物天然产物高效发现[J]. 合成生物学, 2024, 5(3): 492-506

Citation: YU Xuchang, WU Hui, LI Lei. Library construction and targeted BGC screening for more efficient discovery of microbial natural products[J]. Synthetic Biology Journal, 2024, 5(3): 492-506

in high-throughput screening and low-cost DNA sequencing technologies, highly diverse biosynthetic gene clusters (BGCs) have been detected in bacteria and fungi, but characterized compounds are limited, representing the tip of an iceberg, and much more novel small molecules are awaiting for being discovered. Although various strategies have been developed for NP discovery, effectively linking the biosynthetic pathways to their encoded products remains a challenge. Recently, (meta)genomic library construction strategies have shown advantages in elucidating NP biosynthetic pathways more efficiently, and significantly accelerated the discovery of novel NPs by combining with high-efficient targeted BGC screening approaches. In this review, we summarize three strategies for discovering microbial NPs based on (meta)genomic library construction and targeted BGC screening. We also discuss the cloning vectors including Cosmid/Fosmid, BAC/PAC and FAC/YAC, and comment strategies for library construction and targeted BGC screening, such as LEXAS and CONKAT-Seq. Furthermore, we compare strengths, limitations, and applicability of different libraries. At the end, we prospect the future developments of these strategies for the high-throughput discovery of microbial NPs.



**Keywords:** microbial natural products; (meta)genomic mining; biosynthetic gene cluster; library construction; targeted BGC screening

随着全球抗生素耐药性与其他公共健康问题的加剧，新型小分子药物研发迫在眉睫<sup>[1]</sup>。微生物产生大量具有生物活性的天然产物（natural product, NP），是小分子药物的主要来源，在医药、农业与畜牧业等领域发挥重要作用<sup>[2-3]</sup>。鉴于这些天然产物合成相关的基因在微生物基因组中成簇存在，因此对微生物天然产物生物合成基因簇（biosynthetic gene cluster, BGC）的研究是新型天然产物挖掘的重要途径<sup>[4]</sup>。伴随着合成生物学与生物信息学快速发展，大规模微生物（宏）基因组测序揭示了大量新型的功能未知的生物合成基因簇，包括新颖的非核糖体肽（non-ribosomal peptide, NRP）与聚酮（polyketide, PK）等类型合成途径，这为天然产物挖掘提供了一笔宝贵资源<sup>[5-8]</sup>。然而，其中90%以上的天然产物合成基因簇在实验室条件下处于不表达或低表达状态，据统计，目前仅有约3%的微生物来源的天然产物被鉴定，新结构、新靶点化合物严重匮乏<sup>[9-11]</sup>。因此，面对海量的天然产物资源库（功能未知的基因簇），如何从中高效挖掘新结构、新靶点的化合

物仍然是目前研究的重点与难点。

微生物天然产物合成基因簇克隆与异源表达是目前新化合物挖掘的主流策略之一<sup>[12-15]</sup>。一系列不同类型的基因簇直接克隆方法得以建立，包括Gibson组装<sup>[16]</sup>、转化偶联重组（transformation-associated recombination, TAR）<sup>[17]</sup>、Red/ET克隆<sup>[18]</sup>、CRISPR/Cas系统辅助的大片段DNA捕获技术<sup>[19-20]</sup>以及位点特异性重组（site-specific recombination, SSR）介导的克隆<sup>[21]</sup>等。然而，常规的基因簇直接克隆技术较难实现大尺度基因簇（50~150 kb）克隆，其克隆规模也比较有限<sup>[22]</sup>。（宏）基因组文库构建作为一种能够克隆大片段DNA的方法被广泛应用于生物合成基因簇研究<sup>[23-24]</sup>。目前，已经建立的（宏）基因组文库方法类型多样，按照构建载体类型主要包括Cosmid、Fosmid、细菌人工染色体（bacteria artificial chromosome, BAC）、P1人工染色体（P1-derived artificial chromosome, PAC）、真菌人工染色体（fungal artificial chromosome, FAC）与酵母人工染色体（yeast artificial chromosome, YAC）文库等<sup>[25-28]</sup>。由于这些载体系统可有效接受外源片

段, 在克隆超过几百甚至几十万碱基对的大尺度基因簇方面相比直接克隆方法具有明显优势。

然而, 由于文库构建只是随机克隆目标基因簇, 科学家们逐渐开发了多种目标基因簇靶向筛选的策略。最初, 人们借助表型筛选、生物活性检测、原位杂交、HPLC化合物表达筛选等传统筛选方法实现相关基因(簇)的挖掘<sup>[29-30]</sup>。随着生物信息学快速发展, 研究人员逐渐开发出基于全基因组测序技术的天然产物BGC序列引导的筛选与生物信息学功能预测相结合的策略, 并在近年来衍生出各种全新的基于各种大数据库与人工智能等结合的目标基因簇高效筛选方法, 极大促进了海量未知基因簇的挖掘, 加速了新结构天然产物的高效发现。

本文系统阐述了三类(宏)基因组文库构建使用的载体、原理、特征以及构建策略, 并重点介绍了针对大量文库构建所驱动的多种高效基因簇靶向筛选方法的开发及其在不同类型天然产物高效发现中的应用。通过比较不同类型文库的差异及优缺点, 系统总结了基于文库构建与基因簇靶向筛选策略在微生物天然产物挖掘方面的优势, 并展望了这类方法在未来的应用前景。

## 1 Cosmid/Fosmid文库构建驱动微生物天然产物高效挖掘

过去, 人们利用质粒(plasmid)<sup>[31]</sup>和λ噬菌体<sup>[32]</sup>作为两种克隆载体系统引领了DNA片段克隆的时代, 从而为操纵和扩增DNA片段提供了有效的途径。随着分子生物学的发展, 研究人员逐渐发现上述两种常规克隆载体已无法满足大片DNA稳定克隆的需求, 迫切需要开发能容纳更大DNA片段、更加稳定的载体系统。20世纪中后期, Cosmid与Fosmid载体被相继开发用于较大片段DNA的克隆, 这为(宏)基因组文库的构建提供了有效的载体工具。鉴于这两种载体系统均可容纳约40 kb的外源DNA片段, 这极大地提高了克隆完整天然产物合成基因簇的可能性, 从而为新结构天然产物的挖掘提供了有效途径。

### 1.1 Cosmid文库构建及目标基因簇靶向筛选

1978年, Collins等<sup>[33]</sup>构建了一种新型载体Cosmid。该载体系统同时具备λ噬菌体与质粒两种克隆载体的特征, 一方面借助其λ噬菌体DNA *cos*序列, 利用λ噬菌体识别*cos*位点的原理包装外源线性DNA并侵染大肠杆菌, 使DNA在*cos*位点环化形成完整的Cosmid载体; 另一方面借助质粒载体的元件实现DNA环化组装的Cosmid在大肠杆菌体内的复制, 最终实现外源DNA的稳定克隆与复制。鉴于Cosmid载体的主要特征是其克隆的DNA片段大小最高可达40 kb左右, 在克隆完整基因(簇)时具有很好的优势。同时, Cosmid载体具有插入片段分布均匀、遗传稳定性好、拷贝数高等特点<sup>[34]</sup>, 这使得其能够有效地实现大量外源DNA重组片段在大肠杆菌中复制, 从而形成庞大的Cosmid基因(组)文库。

随着大量Cosmid文库的构建, 研究人员开发了许多高效基因簇靶向筛选方法, 并在各种微生物生物功能基因(簇)的研究中得到了广泛应用, 特别是在挖掘克隆天然产物合成基因簇方面。例如, Li等<sup>[35]</sup>通过构建链霉菌TP-A0356菌株Cosmid文库进行基因组挖掘, 使用基因簇序列引导的PCR靶向筛选策略高效鉴定到包含完整链丝菌素(Streptothricin, ST)合成基因簇的Cosmid, 随后通过异源表达发现了化合物Streptothricin F、Streptothricin D以及2个新的ST类似物。杨其会<sup>[36]</sup>采用Cosmid文库构建的方式对链霉菌*Streptomyces netropsis* DSM 40846菌株进行基因组挖掘, 借助基因簇功能预测与BGC序列引导的PCR对行列池和96孔板内含目标基因簇的克隆进行靶向筛选, 成功实现了基因簇的高效鉴定与挖掘, 最后通过异源表达获得了一系列代谢产物, 包括偏端霉素A、Citrulassin B及异呋喃萘醌类似物等活性化合物。张巧燕<sup>[37]</sup>针对新疆特殊生境链霉菌新种*Streptomyces alarensis* TRM 15522构建了Cosmid文库, 并利用序列引导的PCR筛选策略成功挖掘到3个包含较为完整的天然产物合成基因簇的Cosmid, 并通过异源表达差异谱分析发现了新的可能化合物。Gao等<sup>[38]</sup>基于黏细菌*Sorangineae* MSr11367中BGC的功能预测, 通过构建Cosmid基因组文库和对类似

嗜铁素 coelibactin NRPS 基因簇序列进行 PCR 鉴定, 实现了 BGC 靶向筛选与挖掘, 并在 *Myxococcus xanthus* DK1622 中进行异源表达, 最终分离出了一类新颖的嗜铁素 Sorangibactins。

鉴于大部分微生物在实验室条件下不可培养, 宏基因组学的发展为研究这些微生物功能基因提供了有效途径<sup>[39-42]</sup>。基于 Cosmid 构建的微生物宏基因组文库在高效挖掘环境中大量可培养与不可培养微生物所蕴藏的新颖基因簇方面表现出巨大的潜力。例如, Courtois 等<sup>[43]</sup>对土壤环境 DNA 构建了包含 5000 个克隆的 Cosmid 文库, 通过筛选发现了几个新的聚酮合酶合成基因簇并利用异源表达策略获得了其编码的化合物。美国洛克菲勒大学 Brady 教授团队<sup>[44-52]</sup>在土壤宏基因组挖掘方面做出了奠基性工作, 并在基于天然产物序列引导的 PCR 流程的基础上, 结合众多生物信息学数据库与结构预测工具开发了多种高通量基因簇靶向筛选方法, 为新型天然产物的挖掘提供了重要思路与途径。例如, Bauer 等<sup>[46]</sup>利用犹他州土壤的环境 DNA (environment DNA, eDNA) 构建了 Cosmid 文库, 使用一组最小 PKS 特异性简并引物对文库中含有 II 型 PKS BGC 的克隆进行 PCR 靶向筛选, 最后利用异源表达策略在白色链霉菌 J1704 中成功表达了化合物 erdacin 以及两种新的氮杂醌衍生物 utahmycins A 和 B。Libis 等<sup>[51]</sup>构建了土壤宏基因组 Cosmid 文库, 共获得了  $10^7$  个克隆。然后, 他们采用简并引物对 Cosmid 文库池中 NRPS 腺苷酸化结构域或 PKS 酮基合成酶结构域进行了扩增, 获得了一系列天然产物序列标签 (natural products sequence tag, NPST), 并且利用大数据采用基于共现网络的 CONKAT-Seq 策略, 实现了 NRPS 或 PKS 基因簇的精准定位与高通量靶向筛选, 最后通过异源表达在白色链霉菌 J1074 中获得了新化合物 Omnipeptin。Li 等<sup>[52]</sup>也构建了针对土壤的 Cosmid 宏基因组文库, 共获得了超过  $2 \times 10^7$  个克隆。在获得文库中 NRPS 腺苷酸化结构域的 NPST 后, 采用了另一种基于 eSNaPD (environmental Surveyor of Natural Product Diversity) 生物多样性预测软件的谱系分析策略, 实现了在 NRP 数据库中基于最小结构基序的目标基因簇的高通量靶向筛选, 并获得了 3 条新的潜在的甲萘醌结合抗生素

(menaquinone binding antibiotic, MBA) 合成基因簇, 最后利用结构预测化学合成的方式迅速获得了具有显著抑制耐甲氧西林金黄色葡萄球菌 (methicillin-resistant *Staphylococcus aureus*, MRSA) 与多重耐药结核分枝杆菌 (multidrug-resistant *Mycobacterium tuberculosis*) 等活性的化合物 MBA3 (图 1)。

## 1.2 Fosmid 文库构建及目标基因簇靶向筛选

虽然 Cosmid 在 (宏) 基因组文库构建中得到了许多广泛的应用, 但 Cosmid 文库一定程度上存在许多问题。首先 Cosmid 的高拷贝特性使得其容易发生 DNA 重排, 这使克隆的外源片段无法稳定存在<sup>[53]</sup>; 其次由于 Cosmid 文库需要对基因组 DNA 进行酶切从而获得片段化的 DNA, 因此难以克服基因组中酶切位点的选择偏好性, 极大限制了 Cosmid 的应用<sup>[54]</sup>。为解决上述问题, Kim 等<sup>[55]</sup>在 Cosmid 载体 pUCcos 的基础上将携带有大肠杆菌致育因子 (即 F 因子) 的 pBAC 载体与之融合后构建成了一种 pFOS1 的单拷贝 Fosmid 载体。后续经 Epicentre 公司改造, 通过引入诱导型高拷贝复制起始点 *oriV* 获得了 pCC1FOS 和 pCC2FOS 载体<sup>[56]</sup>。由于 Fosmid 载体的单拷贝, 使 DNA 重排概率大大降低, 能够在体内稳定复制, 而诱导型多拷贝又可实现大量载体与克隆的获取<sup>[53, 56]</sup>; 同时, Fosmid 文库是通过机械剪切物理打断获得片段化 DNA, 因此有效地解决了酶切带来的位点选择的偏好性问题<sup>[57]</sup>。鉴于 Fosmid 克隆更加稳定, 且插入片段平均大小与 Cosmid 差不多 (约 40 kb), 因此 Fosmid 逐渐取代 Cosmid 成为一种新的载体系统用于构建 (宏) 基因组文库, 并在各种生物功能基因的研究及天然产物合成基因簇的挖掘方面有着许多的应用。

Felczykowska 等<sup>[58]</sup>构建了 3 个蓝藻的泛基因组文库, 并通过抗菌、抗癌等生物活性评估筛选到了含有潜在功能基因的目标克隆, 并进一步获得了含有活性化合物的提取物。Wolf 等<sup>[59]</sup>针对 *Streptomyces* sp. ATCC 14903 中的生物合成多样性, 构建了一个含有 2880 个克隆的 Fosmid 文库, 并借助序列引导的 PCR 策略对其中的放线酰胺素 actinonin 生物合成基因簇进行了靶向筛选, 最后在

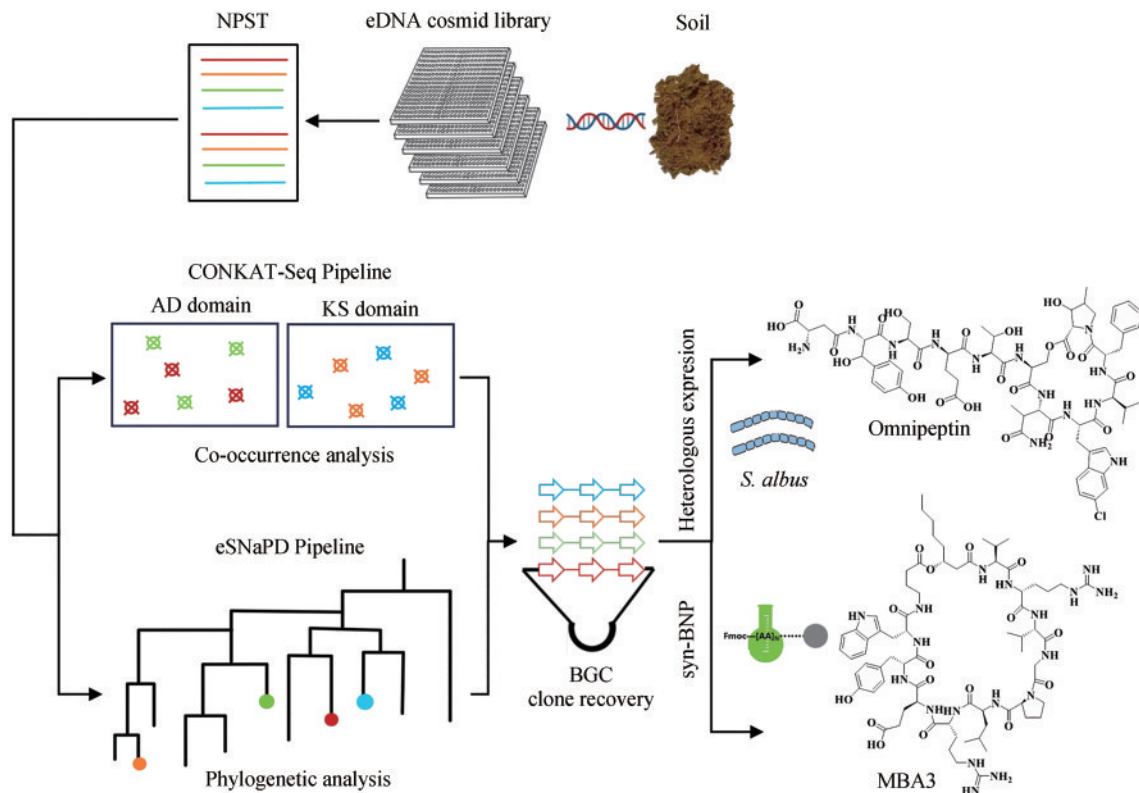


图1 土壤宏基因组 Cosmid文库构建驱动新化合物高效挖掘

(NPST—天然产物序列标签。采用基于共现网络的CONKAT-Seq策略或基于eSNaPD软件的谱系分析获取感兴趣的基因簇，通过异源表达或化学合成方法获取新化合物。)

Fig. 1 Construction of soil metagenomic cosmid libraries for the discovery of novel compounds

(NPST—Natural Products Sequence Tag. The CONKAT-Seq strategy based on co-occurrence network or phylogenetic analysis strategy based on eSNaPD software is used to identify BGCs of interest. Then, novel NPs are obtained by heterologous expression or chemical synthesis.)

白色链霉菌 J1074 异源宿主中成功表达了相应化合物。Jiao 等<sup>[60]</sup> 聚焦 PKS-NRPS 杂合的 *clifednamides* 化合物，构建了 *Kitasatospora* sp. S023 菌株的 Fosmid 文库进行目标基因簇挖掘，使用 PCR 对包含完整 *ct<sub>s023</sub>* 基因簇序列的克隆进行靶向筛选，最后在 *Streptomyces* sp. S001 异源宿主中重构其生物合成途径并成功异源表达了 7 种新的多环四聚体大内酰胺 *clifednamides* D~J 以及两种已知的化合物 *clifednamide* A、*clifednamide* B。

Fosmid 文库也能够实现保存环境样本中所包含的全部微生物基因组 DNA 的目的，并用于后续的基因（簇）功能分析与天然产物挖掘，因此在宏基因组学的研究中也得到广泛应用<sup>[61-62]</sup>。芦晓飞等<sup>[63]</sup> 对西藏米拉山高寒草甸土壤微生物进行 DNA 提取及宏基因组 Fosmid 文库构建，获得了 30624 个克隆，为后续挖掘和研究其中的功能基因奠定了基础。Negri 等<sup>[64]</sup> 构建了包含 83 700 个克隆的土

壤宏基因组 Fosmid 文库，并使用 Nanopore 和 Illumina 测序技术实现重叠群 Contig 组装，利用 antiSMASH 预测 PKS、NRPS、Terpene、Lasso peptide 等 BGC 类型，并结合自主开发的 SNRCM 工具实现高通量靶向筛选鉴定目标 BGC 的目的，为后续天然产物的发现提供了良好的工具。未来，Fosmid 宏基因组文库构建与相应的筛选策略的开发预计将驱动大量环境微生物天然产物的高效发现。

## 2 BAC/PAC 文库构建驱动细菌活性天然产物高效挖掘

事实上，大量微生物天然产物合成基因簇尺度远远大于 40 kb，而 Cosmid/Fosmid 文库在克隆这些基因簇时，受限于克隆尺度往往会出现高度重复序列的克隆，克隆偏好性大大增加，而且造

成叠连序列组装困难，大量 Contig 边界的形成显著提高了目标基因簇筛选的难度。鉴于 Cosmid/Fosmid 文库不能再满足对更大片段 DNA 克隆的要求，为获取大尺度 DNA，克隆载体发展进入以 BAC/PAC 等可插入超过 100 kb 大片段的载体为主的阶段，这也使得 BAC/PAC 基因组文库构建逐渐成为应用更为广泛的大片段完整天然产物合成基因簇挖掘的策略。

## 2.1 BAC 文库构建及目标基因簇靶向筛选

1992 年 Shizuya 等在含有 F 因子的 pMBO131 载体的基础上引入了 *cosN*、*loxP* 等元件构建了第一个 BAC 载体 pBAC108L，后续逐渐引入蓝白斑筛选标记以及其他元件，衍生出了大量的 BAC 系列载体<sup>[65-68]</sup>。由于 BAC 载体含有 RepE、ParA、ParB 等 F 因子元件，因此其具有容纳 100~300 kb 以上外源大片段的能力；BAC 载体由于具有 F 因子单拷贝复制的特征，因此其遗传稳定，保证了外源 DNA 的低缺失、低嵌合与低重组特性<sup>[53]</sup>。总之，BAC 载体逐渐成为近些年来应用最广泛的载体，在克隆大尺度基因簇上具有非常良好的优势。目前基因组 BAC 文库作为广泛应用的 DNA 文库，在不同植物、动物尤其是微生物中得到了广泛应用<sup>[68-79]</sup>。

华中农业大学罗美中教授团队对 BAC 文库的构建方面有着突出的贡献。起初，该团队构建了大量的植物 BAC 文库用于基因组研究<sup>[69]</sup>。2001 年，他们构建了甜瓜耐多病品系 MR-1 的两个 BAC 文库，共获得了 530 个 *Hind* III BAC 克隆和 422 个 *Eco*RI BAC 克隆，并通过探针筛选了抗枯萎病 *Fom-2* 基因用于后续研究<sup>[68]</sup>。而在微生物 BAC 文库构建方面，鉴于链霉菌所展现的巨大生物合成潜力，为建立针对链霉菌的 BAC 文库，罗美中教授团队黄胜等<sup>[70]</sup>构建了用于链霉菌大片段基因组 DNA 克隆与异源表达的 BAC 载体，并利用该载体构建了链霉菌 U27 基因组的 BAC 文库，为链霉菌来源的新型天然产物挖掘提供了一种新的解决方案。借助这个思路，刘家栋等<sup>[71]</sup>为了解阿维链霉菌的生物合成基因簇，解析阿维菌素的生物合成路径，构建了平均插入片段达 101 kb 的 BAC 文库，

获得 2304 个克隆，实现了 25.9 倍基因组的覆盖率。

目前，利用 BAC 文库进行基因簇筛选并获得感兴趣的化合物的策略得到了极其广泛的应用。比如 Geldanamycin、Daptomycin 等临床应用的一线抗生素药物，其生物合成基因簇均是利用 BAC 文库构建的方式通过靶向筛选克隆获取的<sup>[72-73]</sup>。此外，许多具有显著生物活性的微生物天然产物化合物也逐渐利用 BAC 文库构建与基因簇靶向筛选策略得到了挖掘与开发。例如，Liu 等<sup>[74]</sup>构建了针对链霉菌 NRRL 30748 的 BAC 文库，并成功克隆到 Meridamycin 生物合成基因簇，进行了后续的化合物异源表达。Deng 等<sup>[75]</sup>构建了 *S. avermitilis* ATCC 31267 的 BAC 文库，通过 PCR 靶向筛选与异源表达策略，在变铅青链霉菌 1326 菌株中成功产生了 3 种阿维菌素组分。Kazuo Shin-ya 团队<sup>[76]</sup>从小单孢菌 *Micromonospora chalcea* AK-AN57 菌株基因组 BAC 文库中通过目标基因簇序列的上下游 PCR 靶向筛选与异源表达策略在变铅青链霉菌 TK23 中成功获得了化合物 Quinolidomicin；此外，他们还利用 BAC 文库从娄彻氏链霉菌 IFO12908 菌株中通过基因簇序列引导的 PCR 靶向筛选和阿维链霉菌 SUKA32 菌株异源表达策略获得了一种 JBIR-156 新型多烯大环内酰胺化合物<sup>[77]</sup>。上海交通大学陶美凤课题组徐敏等<sup>[78-79]</sup>构建了针对娄彻氏链霉菌 Sal35 的基因组 BAC 文库，平均插入片段达 108 kb，获得了 784 个 BAC 克隆，基因组覆盖率达 10 倍。为了更好地实现目标基因簇的高通量靶向筛选，其团队建立了高通量的文库表达筛选系统 LEXAS，实现了基于“点对点”的变铅青链霉菌 SBT5 的异源表达和抗菌活性筛选，最后对表达的化合物进行结构鉴定，高效获得了疏螺体素、链丝菌素和一种新型羊毛硫肽化合物 Lexapeptide [图 2(a)]。

## 2.2 PAC 文库构建及目标基因簇靶向筛选

随着 P1 噬菌体的发展，Sternberg 等<sup>[80]</sup>基于 P1 噬菌体构建了一种与 Cosmid 工作原理相似的载体，即 P1 噬菌体载体。它含有很多 P1 噬菌体来源的顺式作用元件如 *pac*、*loxP*、Cre 重组酶元件等，可容纳 70~100 kb 大小的 DNA 片段。1994 年，Ioannou 等<sup>[81]</sup>构建了一种由 P1 噬菌体衍生而来的

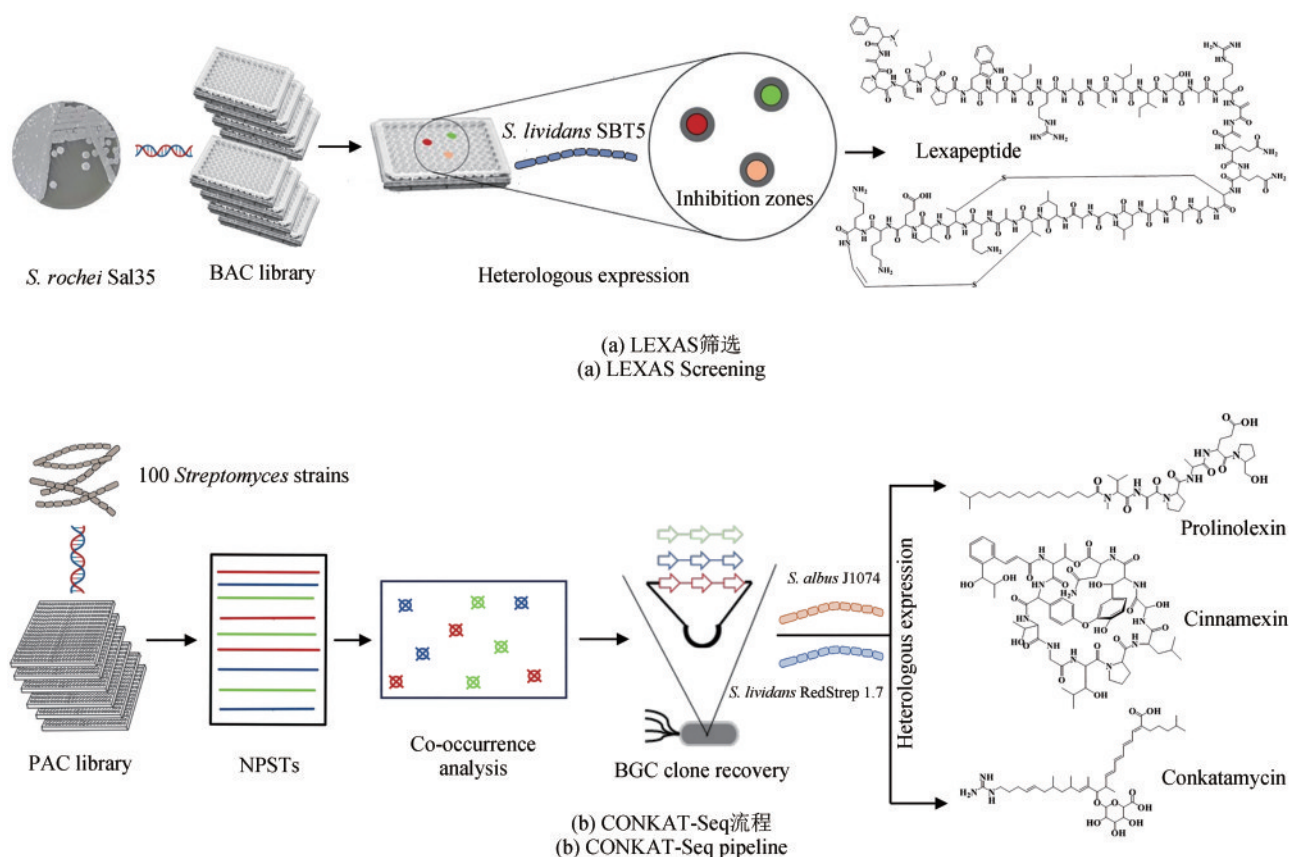


图2 细菌基因组BAC/PAC文库构建加速活性天然产物高通量挖掘。

(a) 构建娄氏链霉菌 *Streptomyces rochei* Sal35 基因组 BAC 文库，建立 LEXAS 筛选系统，高效挖掘新化合物；(b) 构建 100 株链霉菌基因组 PAC 文库，采用基于共现网络的 CONKAT-Seq 策略大规模获取新的非核糖体或聚酮类化合物合成基因簇，并在不同宿主中异源表达，高效获取新化合物

Fig. 2 Construction of bacterial genome BAC/PAC libraries for the high-throughput mining of bioactive natural products.

(a) BAC library construction for the genome of *S. rochei* Sal35 combined with the LEXAS screening system is used for the high-efficient mining of novel compounds; (b) PAC library construction for the genomes of 100 *Streptomyces* combined with the CONKAT-Seq strategy based on co-occurrence network is used to identify novel BGCs encoding NRP and PK, and then heterologous expression can be performed in different hosts to discover novel compounds.

PAC 载体。由于其结合了 P1 噬菌体载体和 F 因子元件的优点，使其同时具备了与 BAC 类似的容纳 100~300 kb 外源 DNA 大片段的能力，同时兼具高稳定性、高转化率等特征。

虽然 PAC 的应用没有 BAC 那么广泛，但目前 PAC 文库在克隆微生物天然产物生物合成基因簇方面也得到了许多应用。例如，Jones 等<sup>[82]</sup> 构建了链霉菌 *Streptomyces tsukubaensis* NRRL 18488 基因组 PAC 文库，共获得了 1920 个 PAC，然后通过针对目标基因簇序列的三对引物 PCR 靶向筛选获得目标基因簇克隆，最后利用异源表达策略在 4 种天蓝色链霉菌衍生菌株中成功表达了重要的免疫抑制剂 FK506 化合物。Castro 等<sup>[83]</sup> 针对来源于超干旱

阿塔卡马沙漠的链霉菌 *Streptomyces leeuwenhoekii* 菌株构建了 PAC 基因组文库进行基因组挖掘，并通过基因簇预测与序列引导的 PCR 靶向筛选策略成功获得了包含目标基因簇的克隆，最后利用异源表达策略成功在天蓝色链霉菌及其衍生菌株中表达了 Chaxamycin 及其衍生物。Tu 等<sup>[84]</sup> 针对海洋链霉菌 *Streptomyces koyangensis* SCSIO 5802 中蕴含的 neoabyssomicin 和 abyssomicin 两种新型化合物生物合成基因簇构建了 PAC 基因组文库，通过基因簇 PCR 靶向筛选与异源表达策略在天蓝色链霉菌 M1152 中成功获得了目标化合物。Libis 等<sup>[85]</sup> 挑选了 100 株链霉菌，构建了含有 60 000 个克隆的 PAC 文库，平均插入片段达 140 kb。他们

针对文库的天然产物合成基因簇采用简并引物扩增获得了NRPS腺苷酸化结构域或PKS酮基合成酶结构域的NPST,采用基于共现网络的CONKAT-Seq筛选策略,高通量快速定位了目标基因簇。最后将含有感兴趣的完整目标基因簇的PAC分别在白色链霉菌J1074和变铅青链霉菌RedStrep 1.7中进行了异源表达,最终获得了Prolinolexin、Cinnamexin和Conkatamycin三种化合物,尤其是Conkatamycin显示出对多重耐药金黄色葡萄球菌的抗菌活性[图2(b)]。

### 3 FAC/YAC文库构建驱动真菌天然产物高效发现

上述基于两大类载体构建的(宏)基因组文库大都建立在以细菌为代表的微生物,而随着微生物天然产物来源研究的不断深入,大量的研究发现丝状真菌基因组中同样蕴藏着巨大的生物合成潜力<sup>[86-89]</sup>。例如,从产黄青霉(*Penicillium chrysogenum*)中发现的青霉素<sup>[90]</sup>、从土曲霉(*Aspergillus terreus*)中发现的洛伐他汀类药物<sup>[91]</sup>、从内生菌内曾发现的紫杉醇<sup>[92]</sup>以及从构巢曲霉(*Aspergillus nidulans*)中发现的棘白菌素类化合物<sup>[93]</sup>均是丝状真菌次级代谢产物的代表。由于丝状真菌遗传背景复杂、操作困难,直接对其基因组进行编辑的难度较大<sup>[94]</sup>。因此基于文库构建的思路克隆编码这些丝状真菌次级代谢产物的生物合成基因簇为挖掘真菌天然产物提供了一种互补的策略。为此,开发出能够携带针对丝状真菌天然产物合成基因簇的载体逐渐成为一种新颖的研究思路,其中新型的FAC载体和应用广泛的YAC载体逐渐被用于构建真菌基因组文库从而获取目标基因簇。

#### 3.1 FAC文库构建及目标基因簇靶向筛选

真菌人工染色体FAC是一种新型的大肠杆菌穿梭真菌人工染色体。2015年,Clevenger等<sup>[95]</sup>基于BAC载体的骨架,通过插入构巢曲霉AMA1真菌自主复制元件,将BAC修饰为FAC,从而构建了FAC载体系统,在大肠杆菌中可稳定克隆100~300 kb的DNA片段。该团队利用这一FAC系统,

构建了不同丝状真菌基因组的FAC文库,发现了包括Terezine D与Valactamide A等多个新型真菌次级代谢产物<sup>[95-96]</sup>。

Bok等<sup>[95]</sup>利用FAC载体系统对丝状真菌*Aspergillus terreus*菌株进行FAC基因组文库的构建,获得了7680个克隆,平均插入大小为100 kb。然后他们通过筛选获得了含有56个基因簇的候选FAC,并挑选了15个FAC在构巢曲霉中进行异源表达,通过LC-HRMS表征了astochrome的生物合成前体化合物Terezine D。Clevenger等<sup>[96]</sup>针对丝状真菌*Aspergillus wentii*、*A. aculeatus*和*A. terreus*构建了包含156个克隆的FAC文库,平均插入片段约100 kb,然后将构建的FAC转入构巢曲霉中进行异源表达。针对发酵代谢产物提取物,采用基于非靶向代谢组学分析和打分系统(FAC-MS与FAC-Score)的高通量筛选策略,对仅在FAC菌株中检测到的化合物给予高分并加以验证,最终获得了Benzomalvin A/D(非核糖体肽)、Sesterterpenoid(萜类)和新化合物Valactamide A(非核糖体肽-聚酮杂合化合物)(图3)<sup>[96]</sup>。

#### 3.2 YAC文库构建及目标基因筛选

酵母人工染色体YAC是一种完全意义上的真菌人工染色体。1983年,Murray等<sup>[97]</sup>仿照酵母染色体构建了具有着丝粒、端粒、自主复制位点等酵母染色体元件的YAC载体,使其能够在酵母中实现稳定复制。YAC在容纳外源DNA片段的能力上具有里程碑式的意义,能实现100~2000 kb片段的插入<sup>[25, 53]</sup>。由于YAC具有接受超大DNA片段的优势,其在常规基因簇克隆上的应用并不多,反而多应用于更大尺度的高等真核生物的复杂基因组文库,比如人类染色体以及一些植物与动物基因组的功能挖掘、基因组物理图谱构建<sup>[98-99]</sup>。

1987年,Burke等<sup>[100]</sup>构建了人类基因组YAC克隆。Saji等<sup>[101]</sup>构建了12条水稻染色体的新型YAC文库。由于YAC插入片段大,只需要少数克隆即可覆盖整个基因组,因此YAC文库常用于进行片段重叠,填补间隙绘制完整的重叠群contig图谱<sup>[53]</sup>。随着微生物天然产物挖掘的不断深入,未

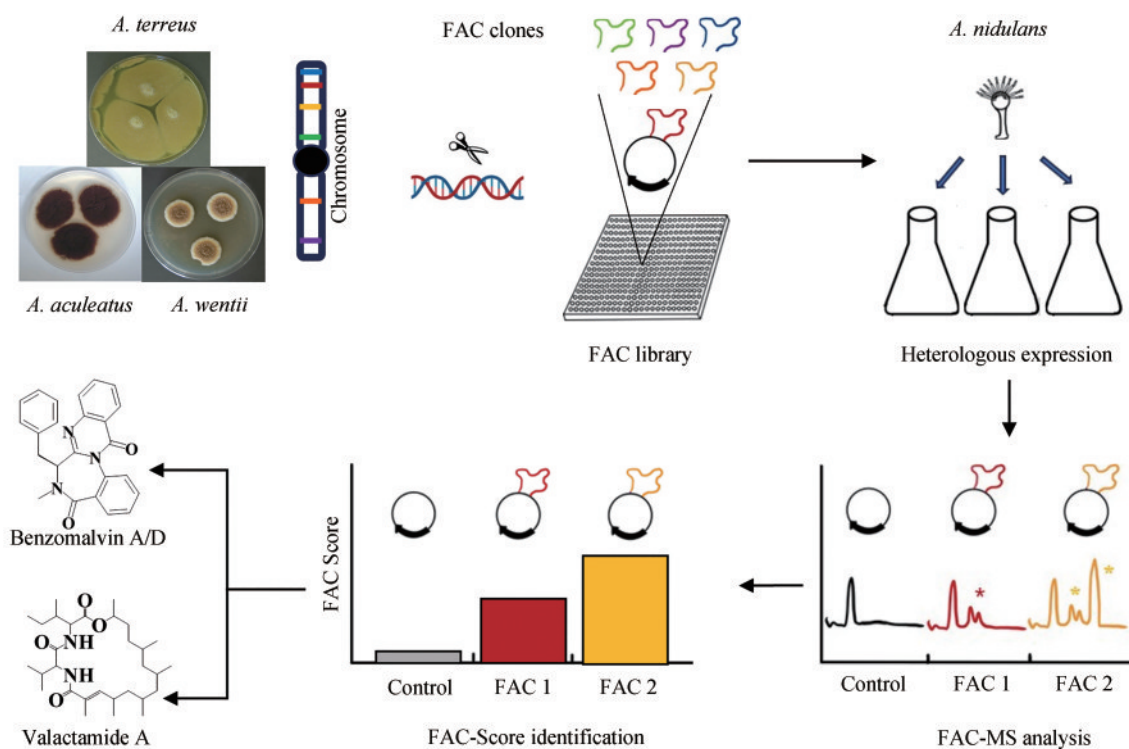


图3 真菌基因组FAC文库构建驱动真菌天然产物创新发现

(采用非靶向代谢组学打分系统, 精准实现新化合物的筛选与分离)

Fig. 3 Construction of fungal genome FAC libraries for discovering fungal natural products

(An untargeted metabolomics scoring system was used to accurately screen and separate novel compounds.)

来可能会发现一些更大尺度的天然产物生物合成基因簇, YAC文库的构建为筛选出含有目标大尺度基因簇提供了可能, 这也为后续进一步挖掘新化合物提供了新途径。

## 4 结论与展望

近年来, 随着大规模微生物(宏)基因组测序技术与生物信息学的快速发展, antiSMASH<sup>[102]</sup>、MIBiG<sup>[103]</sup>、BiG-SCAPE<sup>[104]</sup>或BiG-SLiCE<sup>[105]</sup>等基因组挖掘工具与BGC数据库的开发, 利用(宏)基因组挖掘策略在有效获取新结构天然产物方面展现出了多样化、高通量、规模化的特征。大片段基因簇克隆技术作为天然产物挖掘的一种有效工具逐渐被开发与推广应用。基因组文库构建作为基因簇克隆技术的一种方法, 在最大限度保存基因组信息方面展现出无可比拟的优势, 但大量的筛选工作使研究人员望而生畏。然而, 随着合成生物学、高通量测序与人工智能等生物信息学

技术的发展, 大量的基因簇靶向筛选方法逐渐被开发, 显著加快了微生物天然产物的高效发现<sup>[106]</sup>。本文详细阐述了基于大片段基因簇克隆技术的三种主要(宏)基因组文库类型以及针对基因簇的高效靶向筛选方法, 并介绍了它们在新型微生物天然产物发现中的应用。但目前主流的几种(宏)基因组文库在其载体以及用于新化合物高效发现的策略方面也具有各自的独特优缺点, 详细对比见表1。

Cosmid/Fosmid文库插入片段均匀、稳定性较好, 但插入片段较小, 主要适合于针对eDNA的宏基因组挖掘; BAC/PAC文库具备插入片段较大的特征, 且遗传稳定, 转化效率高, 但由于目前技术限制, 仅适合于单一或有限数量微生物的基因组挖掘, 而针对含有大量eDNA的宏基因组而言, 常常会出现选择性偏差、片段随机性差、文库质量低下、重复率高等问题; FAC/YAC文库在真菌基因组挖掘上具有一定优势, 且克隆片段尺度相比前两种更大, 在基因组物理图谱重构中发挥重要作用, 但克隆稳定性差、嵌合比例高、外源

表1 不同类型文库构建策略用于新化合物高通量发现的比较

文库类型	插入大小	核心元件	筛选方法	文库针对类型	基因簇获取通量	获得化合物	优点	缺点	参考文献
Cosmid	约40 kb	COS位点 质粒元件	CONKAT-seq	宏基因组	10 <sup>7</sup> 个 克隆	Omnipectin(NRP)	适合宏基因组 挖掘	克隆片段大小相 对较小,在克隆大 尺度BGC(>40 kb) 方面存在局限	Libis等 (2019) <sup>[51]</sup>
	约40 kb		eSNaPD	宏基因组	>2×10 <sup>7</sup> 个 克隆	MBA3(NRP)			Li等 (2022) <sup>[52]</sup>
Fosmid	约40 kb	COS位点 F因子元件	序列引导的 PCR筛选	(宏) 基因组	2880个 克隆	Actinonin(NRP)			Wolf等 (2018) <sup>[59]</sup>
BAC	约100 kb	loxP F因子元件	LEXAS	基因组	784个克 隆	Lexapeptide (羊毛硫肽)等	①转化效率高 ②克隆100~ 挖掘	不适合宏基因组	徐敏 (2017) <sup>[78]</sup>
PAC	约140 kb	F因子元件 P1噬菌体 元件	CONKAT-seq	基因组	60 000个 克隆	Prolinolexin(NRP) Cinnamexin(NRP) Conkatamycin(PK)	300 kb片段		Libis等 (2022) <sup>[85]</sup>
FAC	约100 kb	F因子元件 真菌ARS	FAC-MS	基因组	156个 FAC	Benzomalvin A/D (NRP) Sesterterpenoid (Terpene) Valactamide A (Hybrid NRP-PK)	①克隆100~ 300 kb片段 ②适合真菌基 因组挖掘	克隆稳定性差	Clevenger 等(2017) <sup>[96]</sup>
YAC	100~ 2000 kb	着丝粒 端粒 酵母ARS	NA	基因组	1896个 YAC	NA	克隆100~ 2000 kb片段	①克隆稳定性差 ②转化效率低 ③重组率、嵌合 率高	Saji等 (2001) <sup>[101]</sup>

DNA易缺失等问题严重。综上所述,虽然基于不同克隆载体构建的(宏)基因组文库已成功应用于多种天然产物合成基因簇的克隆,但由于不同文库自身存在一定的适用范围,面对着海量功能未知的基因簇,研究者需根据不同目的加以选择。

近年来,随着常规环境来源的微生物中新结构、新靶点化合物发现比例显著下降,研究人员开始转向稀有种类微生物与极端生境或未开发的生境微生物的研究<sup>[107-109]</sup>。这些未知微生物中可能不仅包含了细菌中基因簇丰度最大的链霉菌属<sup>[110]</sup>,也包含了拟无枝酸菌属、库茨纳氏菌属和小单孢菌属等非链霉菌属的稀有放线菌<sup>[111-112]</sup>,以及未鉴定的丝状真菌<sup>[113]</sup>。通过各种天然产物发现方法或将有效揭示其中蕴含的巨大生物潜力。但由于这些微生物大都不可培养<sup>[39-42, 114]</sup>,因此基于文库构建与基因簇靶向筛选策略将极大驱动这些“暗物质”的挖掘,展现出强大的应用价值,预期将为新药创制提供重要化合物资源。

## 参 考 文 献

- [1] HUTCHINGS M I, TRUMAN A W, WILKINSON B. Antibiotics: past, present and future[J]. *Current Opinion in Microbiology*, 2019, 51: 72-80.
- [2] MULLOWNEY M W, DUNCAN K R, ELSAYED S S, et al. Artificial intelligence for natural product drug discovery[J]. *Nature Reviews Drug Discovery*, 2023, 22(11): 895-916.
- [3] ATANASOV A G, ZOTCHEV S B, DIRSCH V M, et al. Natural products in drug discovery: advances and opportunities [J]. *Nature Reviews Drug Discovery*, 2021, 20(3): 200-216.
- [4] KWON M J, STEINIGER C, CAIRNS T C, et al. Beyond the biosynthetic gene cluster paradigm: genome-wide coexpression networks connect clustered and unclustered transcription factors to secondary metabolic pathways[J]. *Microbiology Spectrum*, 2021, 9(2): e00898-21.
- [5] WOOLEY J C, YE Y Z. Metagenomics: facts and artifacts, and computational challenges[J]. *Journal of Computer Science and Technology*, 2010, 25(1): 71-81.
- [6] SIMON C, DANIEL R. Metagenomic analyses: past and future trends[J]. *Applied and Environmental Microbiology*, 2011, 77(4): 1153-1161.
- [7] SCHERLACH K, HERTWECK C. Mining and unearthing hidden biosynthetic potential[J]. *Nature Communications*,

- 2021, 12(1): 3864.
- [8] ZIEMERT N, ALANJARY M, WEBER T. The evolution of genome mining in microbes—a review[J]. *Natural Product Reports*, 2016, 33(8): 988-1005.
- [9] GAVRIILIDOU A, KAUTSAR S A, ZABURANNYI N, et al. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes[J]. *Nature Microbiology*, 2022, 7(5): 726-735.
- [10] BALTZ R H. Gifted microbes for genome mining and natural product discovery[J]. *Journal of Industrial Microbiology & Biotechnology*, 2017, 44(4/5): 573-588.
- [11] LI L. Next-generation synthetic biology approaches for the accelerated discovery of microbial natural products[J]. *Engineering Microbiology*, 2023, 3(1): 100060.
- [12] HUO L J, HUG J J, FU C Z, et al. Heterologous expression of bacterial natural product biosynthetic pathways[J]. *Natural Product Reports*, 2019, 36(10): 1412-1436.
- [13] HAO T T, XIE Z J, WANG M, et al. An anaerobic bacterium host system for heterologous expression of natural product biosynthetic gene clusters[J]. *Nature Communications*, 2019, 10(1): 3665.
- [14] GALM U, SHEN B. Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis[J]. *Expert Opinion on Drug Discovery*, 2006, 1(5): 409-437.
- [15] LI L, JIANG W H, LU Y H. New strategies and approaches for engineering biosynthetic gene clusters of microbial natural products[J]. *Biotechnology Advances*, 2017, 35(8): 936-949.
- [16] GIBSON D G, YOUNG L, CHUANG R Y, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases [J]. *Nature Methods*, 2009, 6(5): 343-345.
- [17] KOUPRINA N, LARIONOV V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology[J]. *Chromosoma*, 2016, 125(4): 621-632.
- [18] ZHANG Y, BUCHHOLZ F, MUYRERS J P, et al. A new logic for DNA engineering using recombination in *Escherichia coli* [J]. *Nature Genetics*, 1998, 20(2): 123-128.
- [19] JIANG W J, ZHAO X J, GABRIELI T, et al. Cas9-assisted targeting of chromosome segments CATCH enables one-step targeted cloning of large gene clusters[J]. *Nature Communications*, 2015, 6: 8101.
- [20] LIANG M D, LIU L S, XU F, et al. Activating cryptic biosynthetic gene cluster through a CRISPR-Cas12a-mediated direct cloning approach[J]. *Nucleic Acids Research*, 2022, 50(6): 3581-3592.
- [21] DU D Y, WANG L, TIAN Y Q, et al. Genome engineering and direct cloning of antibiotic gene clusters *via* phage  $\phi$ BT1 integrase-mediated site-specific recombination in *Streptomyces* [J]. *Scientific Reports*, 2015, 5: 8740.
- [22] 戴岩, 吴旭日, 陈依军. 链霉菌沉默生物合成基因簇激活策略的研究进展[J]. *中国药科大学学报*, 2019, 50(4): 379-388.
- DAI Y, WU X R, CHEN Y J. Advances in strategies for activating silent biosynthetic gene clusters in *Streptomyces*[J]. *Journal of China Pharmaceutical University*, 2019, 50(4): 379-388.
- [23] 戴钊钊. 基于PacBio的高通量Fosmid文库克隆长配末端测序技术的开发[D]. 武汉: 华中农业大学, 2019.
- DAI Z Z. High-throughput long paired-end sequencing of a Fosmid library by PacBio[D]. Wuhan: Huazhong Agricultural University, 2019.
- [24] PETERSON D G, TOMKINS J P, FRISCH D A, et al. Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide[J/OL]. *Journal of Agricultural Genomics*, 2000, 5: 1-100[2023-12-01]. <https://www.mgsl.msstate.edu/pubs/Peterson%20et%20al%202000.pdf>.
- [25] 樊颖伦, 赵开军. 大片段克隆载体研究进展[J]. *中国生物工程杂志*, 2004, 24(3): 12-16.
- FAN Y L, ZHAO K J. Progress in development of large DNA fragment cloning vectors[J]. *China Biotechnology*, 2004, 24(3): 12-16.
- [26] HOSODA F, NISHIMURA S, UCHIDA H, et al. An F factor based cloning system for large DNA fragments[J]. *Nucleic Acids Research*, 1990, 18(13): 3863-3869.
- [27] DANIEL R. The metagenomics of soil[J]. *Nature Reviews Microbiology*, 2005, 3(6): 470-478.
- [28] 高花花, 胡娟, 刘玲丽. 基因簇大片段克隆技术研究进展及挑战[J]. *微生物学通报*, 2023, 50(1): 351-367.
- GAO H H, HU J, LIU L L. Research progress and challenges in cloning large gene clusters[J]. *Microbiology China*, 2023, 50(1): 351-367.
- [29] 王凯, 张燕洁, 关兵, 等. 棉花细菌人工染色体的荧光原位杂交(BAC-FISH)技术[J]. *生物化学与生物物理进展*, 2007, 34(11): 1216-1222.
- WANG K, ZHANG Y J, GUAN B, et al. Fluorescence *in situ* hybridization of bacterial artificial chromosome in cotton[J]. *Progress in Biochemistry and Biophysics*, 2007, 34(11): 1216-1222.
- [30] 池溟甜, 陈实. 基因组学技术解码天然产物合成[J]. *生物工程学报*, 2019, 35(10): 1889-1900.
- CHI H T, CHEN S. Genomics approaches decode natural products synthesis[J]. *Chinese Journal of Biotechnology*, 2019, 35(10): 1889-1900.
- [31] SUMMERS W C. Plasmids: histories of a concept[M/OL]// *Reticulate evolution: symbiogenesis, lateral gene transfer, hybridization and infectious heredity*. Cham: Springer, 2015: 179-190[2023-12-01]. [https://link.springer.com/chapter/10.1007/978-3-319-16345-1\\_6](https://link.springer.com/chapter/10.1007/978-3-319-16345-1_6).
- [32] LEDERBERG E M, LEDERBERG J. Genetic studies of

- lysogenicity in *Escherichia coli*[J]. *Genetics*, 1953, 38(1): 51-64.
- [33] COLLINS J, HOHN B. Cosmids: a type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage lambda heads[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1978, 75(9): 4242-4246.
- [34] WENZEL R, HERRMANN R. Cosmid cloning with small genomes[M/OL]//Nonmammalian genomic analysis. San Diego, CA, USA: Academic Press, 1996: 197-222[2023-12-01]. <https://www.sciencedirect.com/science/article/abs/pii/B9780121012854500084?via%3Dihub>.
- [35] LI J E, GUO Z Y, HUANG W, et al. Mining of a streptothricin gene cluster from *Streptomyces* sp. TP-A0356 genome via heterologous expression[J]. *Science China Life Sciences*, 2013, 56(7): 619-627.
- [36] 杨其会. 基于基因组挖掘策略研究 *Streptomyces netropsis* DSM 40846 中天然产物生物合成基因簇[D]. 武汉: 华中农业大学, 2022.
- YANG Q H. Exploring natural product biosynthesis gene clusters in *Streptomyces netropsis* DSM 40846 by genome mining[D]. Wuhan: Huazhong Agricultural University, 2022.
- [37] 张巧燕. 基于 genome mining 技术的 TRM 15522 次级代谢产物研究[D]. 阿拉尔: 塔里木大学, 2021.
- ZHANG Q Y. Study on secondary metabolites of TRM 15522 based on genome mining technology[D]. Alaer: Tarim University, 2021.
- [38] GAO Y S, WALT C, BADER C D, et al. Genome-guided discovery of the myxobacterial thiolactone-containing sorangibactins[J]. *ACS Chemical Biology*, 2023, 18(4): 924-932.
- [39] LIU J Y, YANG L, KJELLERUP B V, et al. Viable but nonculturable (VBNC) state, an underestimated and controversial microbial survival strategy[J]. *Trends in Microbiology*, 2023, 31(10): 1013-1023.
- [40] PINTO D, SANTOS M A, CHAMBEL L. Thirty years of viable but nonculturable state research: unsolved molecular mechanisms[J]. *Critical Reviews in Microbiology*, 2015, 41(1): 61-76.
- [41] EKKERS D M, CRETOIU M S, KIELAK A M, et al. The great screen anomaly: a new frontier in product discovery through functional metagenomics[J]. *Applied Microbiology and Biotechnology*, 2012, 93(3): 1005-1020.
- [42] HANDELSMAN J. Metagenomics: application of genomics to uncultured microorganisms[J]. *Microbiology and Molecular Biology Reviews*, 2004, 68(4): 669-685.
- [43] COURTOIS S, CAPPELLANO C M, BALL M, et al. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products [J]. *Applied and Environmental Microbiology*, 2003, 69(1): 49-55.
- [44] BRADY S F, CHAO C J, HANDELSMAN J, et al. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA[J]. *Organic Letters*, 2001, 3(13): 1981-1984.
- [45] BRADY S F, CLARDY J. Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA[J]. *Angewandte Chemie International Edition*, 2005, 44(25): 7225-7227.
- [46] BAUER J D, KING R W, BRADY S F. Utahmycins A and B, azaquinones produced by an environmental DNA clone[J]. *Journal of Natural Products*, 2010, 73(5): 976-979.
- [47] KALLIFIDAS D, KANG H S, BRADY S F. Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters[J]. *Journal of the American Chemical Society*, 2012, 134(48): 19552-19555.
- [48] KANG H S, BRADY S F. Arimetamycin A: improving clinically relevant families of natural products through sequence-guided screening of soil metagenomes[J]. *Angewandte Chemie International Edition*, 2013, 52(42): 11063-11067.
- [49] PEEK J, LILIC M, MONTIEL D, et al. Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism[J]. *Nature Communications*, 2018, 9(1): 4147.
- [50] HOVER B M, KIM S H, KATZ M, et al. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens[J]. *Nature Microbiology*, 2018, 3(4): 415-422.
- [51] LIBIS V, ANTONOVSKY N, ZHANG M Y, et al. Uncovering the biosynthetic potential of rare metagenomic DNA using co-occurrence network analysis of targeted sequences[J]. *Nature Communications*, 2019, 10(1): 3848.
- [52] LI L, KOIRALA B, HERNANDEZ Y, et al. Identification of structurally diverse menaquinone-binding antibiotics with *in vivo* activity against multidrug-resistant pathogens[J]. *Nature Microbiology*, 2022, 7(1): 120-131.
- [53] 刘建喜, 林爱星, 丁翔, 等. DNA 克隆载体的发展和应[J]. *遗传*, 1997, 19(6): 41-44.
- LIU J X, LIN A X, DING X, et al. The development and application of DNA cloning vectors[J]. *Hereditas(Beijing)*, 1997, 19(6): 41-44.
- [54] EVANS G A, LEWIS K A. Physical mapping of complex genomes by cosmid multiplex analysis[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1989, 86(13): 5030-5034.
- [55] KIM U J, SHIZUYA H, DE JONG P J, et al. Stable propagation of cosmid sized human DNA inserts in an F factor

- based vector[J]. *Nucleic Acids Research*, 1992, 20(5): 1083-1085.
- [56] 李昂, 张安, 唐君, 等. Fosmid 基因组文库构建及应用现状[J]. *江苏农业科学*, 2013, 41(6): 28-30.  
LI A, ZHANG A, TANG J, et al. Fosmid genomic library construction and its application status[J]. *Jiangsu Agricultural Sciences*, 2013, 41(6): 28-30.
- [57] LIU C X, LIU X L, LEI L, et al. Fosmid library construction and screening for the maize mutant gene *Vestigial glume 1*[J]. *The Crop Journal*, 2016, 4(1): 55-60.
- [58] FELCZYKOWSKA A, DYDECKA A, BOHDANOWICZ M, et al. The use of fosmid metagenomic libraries in preliminary screening for various biological activities[J]. *Microbial Cell Factories*, 2014, 13: 105.
- [59] WOLF F, LEIPOLDT F, KULIK A, et al. Characterization of the actinonin biosynthetic gene cluster[J]. *ChemBioChem*, 2018, 19(11): 1189-1195.
- [60] JIAO Y J, LIU Y, WANG H X, et al. Expression of the clifednamide biosynthetic pathway in *Streptomyces* generates 27, 28-*seco*-derivatives[J]. *Journal of Natural Products*, 2020, 83(9): 2803-2808.
- [61] UFARTÉ L, BOZONNET S, LAVILLE E, et al. Functional metagenomics: construction and high-throughput screening of Fosmid libraries for discovery of novel carbohydrate-active enzymes[M/OL]//*Methods in molecular biology: microbial environmental genomics*. New York: Humana Press, 2016, 1399: 257-271[2023-12-01]. [https://link.springer.com/protocol/10.1007/978-1-4939-3369-3\\_15](https://link.springer.com/protocol/10.1007/978-1-4939-3369-3_15).
- [62] RIESENFELD C S, SCHLOSS P D, HANDELSMAN J. Metagenomics: genomic analysis of microbial communities[J]. *Annual Review of Genetics*, 2004, 38: 525-552.
- [63] 芦晓飞, 赵志祥, 谢丙炎, 等. 西藏米拉山高寒草甸土壤微生物 DNA 提取及宏基因组 Fosmid 文库构建[J]. *应用与环境微生物学报*, 2009, 15(6): 824-829.  
LU X F, ZHAO Z X, XIE B Y, et al. DNA extraction and construction of a metagenomic Fosmid library of alpine meadow soil from the Mila mountains in Tibet, China[J]. *Chinese Journal of Applied and Environmental Biology*, 2009, 15(6): 824-829.
- [64] NEGRI T, MANTRI S, ANGELOV A, et al. A rapid and efficient strategy to identify and recover biosynthetic gene clusters from soil metagenomes[J]. *Applied Microbiology and Biotechnology*, 2022, 106(8): 3293-3306.
- [65] SHIZUYA H, BIRREN B, KIM U J, et al. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1992, 89(18): 8794-8797.
- [66] KIM U J, BIRREN B W, SLEPAK T, et al. Construction and characterization of a human bacterial artificial chromosome library[J]. *Genomics*, 1996, 34(2): 213-218.
- [67] LIU Y G, SHIRANO Y, FUKAKI H, et al. Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, 96(11): 6535-6540.
- [68] LUO M Z, WANG Y H, FRISCH D, et al. Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium* wilt (Fom-2)[J]. *Genome*, 2001, 44(2): 154-162.
- [69] LUO M Z, WING R A. An improved method for plant BAC library construction[M/OL]//*Methods in molecular biology: plant functional genomics*. New York: Human Press, 2003, 236: 3-20[2023-12-01]. <https://link.springer.com/protocol/10.1385/1-59259-413-1:3>.
- [70] 黄胜, 李娜, 周俊, 等. 适用于链霉菌大片段基因组 DNA 克隆和异源表达的细菌人工染色体(BAC)载体的构建及应用[J]. *微生物学报*, 2012, 52(1): 30-37.  
HUANG S, LI N, ZHOU J, et al. Construction of a new bacterial artificial chromosome (BAC) vector for cloning of large DNA fragments and heterologous expression in *Streptomyces*[J]. *Acta Microbiologica Sinica*, 2012, 52(1): 30-37.
- [71] 刘家栋, 王革娇, 罗美中. 阿维链霉菌 BAC 文库的构建及分析[J]. *华中农业大学学报*, 2016, 35(5): 45-50.  
LIU J D, WANG G J, LUO M Z. Construction and analysis of a BAC library of *Streptomyces avermitilis* genome[J]. *Journal of Huazhong Agricultural University*, 2016, 35(5): 45-50.
- [72] RASCHER A, HU Z H, VISWANATHAN N, et al. Cloning and characterization of a gene cluster for geldanamycin production in *Streptomyces hygroscopicus* NRRL 3602[J]. *FEMS Microbiology Letters*, 2003, 218(2): 223-230.
- [73] PENN J, LI X, WHITING A, et al. Heterologous production of daptomycin in *Streptomyces lividans*[J]. *Journal of Industrial Microbiology & Biotechnology*, 2006, 33(2): 121-128.
- [74] LIU H B, JIANG H, HALTLI B, et al. Rapid cloning and heterologous expression of the meridamycin biosynthetic gene cluster using a versatile *Escherichia coli*-*Streptomyces* artificial chromosome vector, pSBAC[J]. *Journal of Natural Products*, 2009, 72(3): 389-395.
- [75] DENG Q, ZHOU L, LUO M Z, et al. Heterologous expression of avermectins biosynthetic gene cluster by construction of a Bacterial Artificial Chromosome library of the producers[J]. *Synthetic and Systems Biotechnology*, 2017, 2(1): 59-64.
- [76] HASHIMOTO T, HASHIMOTO J, KOZONE I, et al. Biosynthesis of quinolidomicin, the largest known macrolide

- of terrestrial origin: identification and heterologous expression of a biosynthetic gene cluster over 200 kb[J]. *Organic Letters*, 2018, 20(24): 7996-7999.
- [77] HASHIMOTO T, KOZONE I, HASHIMOTO J, et al. Novel macrolactam compound produced by the heterologous expression of a large cryptic biosynthetic gene cluster of *Streptomyces rochei* IFO12908[J]. *The Journal of Antibiotics*, 2020, 73(3): 171-174.
- [78] 徐敏. 高通量异源表达辅助的活性天然产物挖掘与新型羊毛硫肽 Lexapeptide 的生物合成研究[D]. 上海: 上海交通大学, 2017. XU M. Discovery of bioactive natural products enabled by high throughput heterologous expression and biosynthesis of a novel lantibiotic, lexapeptide[D]. Shanghai: Shanghai Jiao Tong University, 2017.
- [79] XU M, ZHANG F, CHENG Z, et al. Functional genome mining reveals a class V lanthipeptide containing a D-amino acid introduced by an F<sub>420</sub> H<sub>2</sub>-dependent reductase[J]. *Angewandte Chemie International Edition*, 2020, 59(41): 18029-18035.
- [80] STERNBERG N. Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1990, 87(1): 103-107.
- [81] IOANNOU P A, AMEMIYA C T, GARNES J, et al. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments[J]. *Nature Genetics*, 1994, 6(1): 84-89.
- [82] JONES A C, GUST B, KULIK A, et al. Phage p1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster[J]. *PLoS One*, 2013, 8(7): e69319.
- [83] CASTRO J F, RAZMILIC V, GOMEZ-ESCRIBANO J P, et al. Identification and heterologous expression of the chaxamycin biosynthesis gene cluster from *Streptomyces leeuwenhoekii*[J]. *Applied and Environmental Microbiology*, 2015, 81(17): 5820-5831.
- [84] TU J J, LI S T, CHEN J, et al. Characterization and heterologous expression of the neoabyssomicin/abyssomicin biosynthetic gene cluster from *Streptomyces koyangensis* SCSIO 5802[J]. *Microbial Cell Factories*, 2018, 17(1): 28.
- [85] LIBIS V, MACINTYRE L W, MEHMOOD R, et al. Multiplexed mobilization and expression of biosynthetic gene clusters[J]. *Nature Communications*, 2022, 13(1): 5256.
- [86] BÉRDY J. Bioactive microbial metabolites[J]. *The Journal of Antibiotics*, 2005, 58(1): 1-26.
- [87] CHIANG C Y, OHASHI M, TANG Y. Deciphering chemical logic of fungal natural product biosynthesis through heterologous expression and genome mining[J]. *Natural Product Reports*, 2023, 40(1): 89-127.
- [88] KELLER N P. Translating biosynthetic gene clusters into fungal armor and weaponry[J]. *Nature Chemical Biology*, 2015, 11(9): 671-677.
- [89] ROKAS A, WISECAVER J H, LIND A L. The birth, evolution and death of metabolic gene clusters in fungi[J]. *Nature Reviews Microbiology*, 2018, 16(12): 731-744.
- [90] VAN DEN BERG M A, ALBANG R, ALBERMANN K, et al. Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*[J]. *Nature Biotechnology*, 2008, 26(10): 1161-1168.
- [91] ALBERTS A W, CHEN J, KURON G, et al. Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1980, 77(7): 3957-3961.
- [92] STIERLE A, STROBEL G, STIERLE D. Taxol and taxane production by *Taxomyces andreanae*, an endophytic fungus of Pacific yew[J]. *Science*, 1993, 260(5105): 214-216.
- [93] HÜTTEL W. Echinocandins: structural diversity, biosynthesis, and development of antimycotics[J]. *Applied Microbiology and Biotechnology*, 2021, 105(1): 55-66.
- [94] 林继聪, 邹根, 刘宏民, 等. CRISPR/Cas 基因组编辑技术在丝状真菌次级代谢产物合成中的应用[J]. *合成生物学*, 2023, 4(4): 738-755. LIN J C, ZOU G, LIU H M, et al. Application of CRISPR/Cas genome editing technology in the synthesis of secondary metabolites of filamentous fungi[J]. *Synthetic Biology Journal*, 2023, 4(4): 738-755.
- [95] BOK J W, YE R, CLEVINGER K D, et al. Fungal artificial chromosomes for mining of the fungal secondary metabolome [J]. *BMC Genomics*, 2015, 16(1): 343.
- [96] CLEVINGER K D, BOK J W, YE R, et al. A scalable platform to identify fungal secondary metabolites and their gene clusters [J]. *Nature Chemical Biology*, 2017, 13(8): 895-901.
- [97] MURRAY A W, SZOSTAK J W. Construction of artificial chromosomes in yeast[J]. *Nature*, 1983, 305(5931): 189-193.
- [98] SANTRA D K, SANDHU D, TAI T, et al. Construction and characterization of a soybean yeast artificial chromosome library and identification of clones for the Rps6 region[J]. *Functional & Integrative Genomics*, 2003, 3(4): 153-159.
- [99] KURATA N, UMEHARA Y, TANOUE H, et al. Physical mapping of the rice genome with YAC clones[J]. *Plant Molecular Biology*, 1997, 35(1/2): 101-113.
- [100] BURKE D T, CARLE G F, OLSON M V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors[J]. *Science*, 1987, 236(4803): 806-812.
- [101] SAJI S, UMEHARA Y, ANTONIO B A, et al. A physical map with yeast artificial chromosome (YAC) clones covering 63% of the 12 rice chromosomes[J]. *Genome*, 2001, 44(1): 32-37.
- [102] BLIN K, SHAW S, KLOOSTERMAN A M, et al. antiSMASH

- 6.0: improving cluster detection and comparison capabilities [J]. *Nucleic Acids Research*, 2021, 49(W1): W29-W35.
- [103] TERLOUW B R, BLIN K, NAVARRO-MUÑOZ J C, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters[J]. *Nucleic Acids Research*, 2023, 51(D1): D603-D610.
- [104] NAVARRO-MUÑOZ J C, SELEM-MOJICA N, MULLOWNEY M W, et al. A computational framework to explore large-scale biosynthetic diversity[J]. *Nature Chemical Biology*, 2020, 16 (1): 60-68.
- [105] KAUTSAR S A, VAN DER HOOFT J J J, RIDDER D D, et al. BiG-SLiCE: a highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters[J]. *GigaScience*, 2021, 10 (1): g1aa154.
- [106] 赖奇龙, 姚帅, 查毓国, 等. 微生物组生物合成基因簇发掘方法及应用前景[J]. *合成生物学*, 2023, 4(3): 611-627.
- LAI Q L, YAO S, ZHA Y G, et al. Microbiome-based biosynthetic gene cluster data mining techniques and application potentials[J]. *Synthetic Biology Journal*, 2023, 4 (3): 611-627.
- [107] LI L. Accessing hidden microbial biosynthetic potential from underexplored sources for novel drug discovery[J]. *Biotechnology Advances*, 2023, 66: 108176.
- [108] BALTZ R H. Marcel Faber Roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? [J]. *Journal of Industrial Microbiology & Biotechnology*, 2006, 33(7): 507-513.
- [109] SHU W S, HUANG L N. Microbial diversity in extreme environments[J]. *Nature Reviews Microbiology*, 2022, 20(4): 219-235.
- [110] LI S, DONG L, LIAN W H, et al. Exploring untapped potential of *Streptomyces* spp. in Gurbantunggut Desert by use of highly selective culture strategy[J]. *The Science of the Total Environment*, 2021, 790: 148235.
- [111] LAZZARINI A, CAVALETTI L, TOPPO G, et al. Rare Genera of actinomycetes as potential producers of new antibiotics[J]. *Antonie Van Leeuwenhoek*, 2000, 78(3/4): 399-405.
- [112] XIE F Y, PATHOM-AREE W. Actinobacteria from desert: diversity and biotechnological applications[J]. *Frontiers in Microbiology*, 2021, 12: 765531.
- [113] WANASINGHE D N, MORTIMER P E, XU J C. Insight into the systematics of microfungi colonizing dead woody twigs of *Dodonaea viscosa* in Honghe (China) [J]. *Journal of Fungi*, 2021, 7(3): 180.
- [114] LEWIS K, EPSTEIN S, D'ONOFRIO A, et al. Uncultured microorganisms as a source of secondary metabolites[J]. *The Journal of Antibiotics*, 2010, 63(8): 468-476.



**通讯作者:** 李雷(1989—), 男, 长聘教轨副教授, 博士生导师。研究方向为微生物天然药物化学与合成生物学。  
E-mail: lei.li@sjtu.edu.cn



**通讯作者:** 吴辉(1982—), 男, 教授, 博士生导师。研究方向为微生物代谢设计与重编程。  
E-mail: hwu@ecust.edu.cn



**第一作者:** 虞旭昶(2000—), 男, 硕士研究生。研究方向为微生物天然产物创新发现与高效制造。  
E-mail: yxc@ecust.edu.cn